Temporal alignment of communicative gesture sequences

Alexis Heloir[*], Nicolas Courty[*], Sylvie Gibet[*], Franck Multon[†]

[*]Samsara, UBS, Bat. Yves Coppens, BP 573, 56017 Vannes cedex, F

[†] LPBEM, University Rennes 2, av. Charles Tillon, 35044 Rennes, F

[†] SIAMES project, IRISA, Campus de Beaulieu, 35042 Rennes, F

email: [*]firstname.surname@univ-ubs.fr, [†]fmulton@irisa.fr

Abstract

In this paper we address the problem of temporal alignment applied to captured communicative gestures conveying different styles. We propose a representation space that may be considered as *robust* to the spatial variability induced by style. By extending a multilevel dynamic time warping algorithm, we show how this extension can fulfil the goals of time correspondence between gesture sequences while preventing jerkiness introduced by standard time warping methods.

Keywords: Animation, conversational agents, gestural communication, style translation

Introduction

Achievement of a virtual humanoid capable of performing realistic and pleasant movements like the gestures involved in sign language communication is still a challenge within the animation community.

In a conversational situation, style conveys useful hints to verbal and nonverbal features of the discourse such as nuances, intensity, emphasis points, speaker genre, cultural background, and emotional state. Consequently, automatic generation of expressive human motion requires methods that are capable of seamlessly handling a wide range of different styles along the animation.

Many recent works provide new insights into motion style, and rely on motion capture data to identify various components of motion. These works may provide animators with tools to interactively edit and manipulate motion, or methods which can be integrated into data-driven animation frameworks. Such approaches describe motion as the combination of components representing respectively both the content and the style. Our approach is in the line of these works but focuses on communicative gestures.

We will now introduce the following definition of gestural style on which we will rely during the rest of this paper: we consider that style is the variability observed among two realizations of the same gestural sequence. This definition is volontarily low level, signal oriented as our investigations are motivated by motion signal analysis.

We worked on multiple realizations of a sequence of French Sign Language (FSL) ges-

tures. To do so, we asked a professional signer to perform several motion capture records of a predefined FSL sequence by varying several aspects of the discourse: mood, emphasis and speed.

Dealing with such data raises many difficulties. One of the most obvious is that a communicative gestural sequence (CGS) is by nature non periodic. A consequence of this particular limitation is that CGS falls out of the range of application of many existing methods that have proved to work over pseudo periodic motions like locomotion. On the other hand, CGS elementary gestures are tightly linked together through a coarticulation[1] process which challenges both manual and automatic segmentation, especially when handling multiple styles. As illustrated in figure 1, critical aspects of the influence of style on CGS are both of temporal and spacial aspects.

In this work, we aim at capturing the temporal features of several realizations of a CGS. We rely on the assumption that there exists a *fundamental motion* which is common to multiple styled realizations of a CGS exists. Then we show that the Weighted PCA representation space is compatible with the *fundamental motion* assumption. By emphasis this preliminary result, we then propose a multi level dynamic time warping (DTW) algorithm that is well suited for the problem of CGS alignment by resolving both local and global timing variances induced by style while preventing jerkiness and being robust to artifacts introduced by the signer. We then illustrate our work by performing a temporal alignment between a

[1]Coarticulation is manifested by the fact that a motion primitive is highly influenced by the previous and the following primitives in a CGS

realization of a CGS performed in an *as neutral as possible* manner and other realizations of this CGS performed according to different moods, emphasis and execution speeds.

The rest of this paper is organized as follows: in the first section, we present the existing works that address the problem of styled motion edition, then we introduce the style robust distance metric on which The Adaptive Fast Time Warping algorithm introduced in section 3 steps on. Section 4 introduces the sign language CGS that we have captured for the experiments detailed in section 5. Results are discussed in section 6. We then conclude by drawing perspectives of this work.

Related works

In order to introduce the problem of style retrieval and analysis from captured motions, we first show how model-based animation methods have somehow failed to produce yet natural and convincing motions. We then present works concerned by the characterization of style in motion sequences, and then give some insights on the temporal alignment problem between two motions.

Model-based animation methods Studies on sign language which have been carried out since the 60's have lead to dedicated description/transcription systems [1]. Several gestural generation systems inspired by this paradigm have appeared since then. Lebourque & al. [2] propose an expressive gesture synthesis system where the task is expressed as a discrete

sequence of targets in the euclidian space around the virtual signer. The sequence of postures is then solved thanks to a sensory-motor inverse kinematics solver. More recently, the ESIGN project [3] designed and set up a communicative gesture synthesis system driven by sign language transcriptions.

More style-centric studies around human motion appeared by the end of the XIX^e century and have been continued and enhanced until today [4]. The underlying theory derived from those works served as a base for procedural motion synthesis systems [5]. Other procedural systems step on psycho cognitive studies [6] in order to convey stylistic features [7].

Procedural systems have proved to be capable of providing understandable expressive gestures with a great level of control. However, such generative models, by relying on kinematics models, have failed to produce natural, smooth motions. This failure was partially solved by intensive use of captured motions. As a consequence, many recent works related to human motion style rely on motion capture.

Data driven style characterization characterization and extraction of the style for motion captured data is tightly linked to motion editing methods. The first works which focus on this particular aspect are inspired by signal processing. Unuma & al. [8] apply Fourier decomposition to motion signal and identify style features like briskness or the amount of weariness in the motion. Bruderlin & al. [9] perform frequency decomposition of motion signal by using multiresolution filtering. Motion blending is then obtained by separately handling motion frequency components. Amaya & al introduced the notion of *emotion*

between a *neutral* motion and an *emotionally charged* motion [10]. The method is able to endow an arbitrary *neutral motion* with the emotion embeded in the initial *emotional motion*. Several methods based on captured data exploit structured bases of motion segments [11, 12, 13]. By finding the best sequence of segments matching user-defined tasks or constraints, the system is able to produce arbitrary long motions. However, those methods do not directly address style editing, as the style of generated motion is bounded to the inner style of the captured motion.

In parallel, style caracterization has motivated the use of statistical methods like independant component analysis (ICA) [14, 15], PCA [16] or factorization models [17, 18]. Other works are inspired by generative approaches based either on structure dicovery and mapping of discrete states embedded in two styled motions [19, 20]. While these methods reveals to efficiently perform spacial style translation of either body or facial motion, they do not directly address the temporal variations induced by style.

Temporal characterization of style It has been well accepted that style affects both temporal and spacial characteristics of human motion. As far as we know, those two concerns have been addressed in a separate manner. Temporal and spacial aspects of style are characterized separately and recombined together during the motion generation step. Characterization of temporal stylistic features is by itself a non trivial problem and has been addressed, in most cases by relying on a well known non linear time alignment method: dynamic time warping (DTW) [21]. Bruderlin [9] adopted Sederberg's shape based algorithm to vectors

of postures described as Euler angles. Witkin & al. [22] adopted a motion curve based warping scheme over euler-angular posture vectors confining the motion editing process to a set of chosen articular trajectories. Rose [23] performs time warping over motion curves described as uniform cubic B-spline curves defined by their control points. Gleicher [24] adopts constrained dynamic time warping to align motions to perform relevant blending. Hsu & al. [25] propose an iterative time warping algorithm which directly operates on posture vectors. Most recently, Shapiro [15] proposes an interactive motion editing tool that lets the user choose a relevant joint defined in Euclidean coordinates to perform time alignment over two styled motions.

Forbes & al. [26] introduce a motion search algorithm based on a *weighted PCA-based* pose representation. This algorithm evaluates a time-warp distance between sequences of postures by performing bi-directional DTW from a seed point in a distance matrix built over the *PCA-based* representation space. The *weighted PCA-based* representation space fits well with our need of splitting a CGS into two parts: the *fundamental motion* that conveys the meaningful part of a gesture sequence and the *stylistic content* of the motion wich conveys stylistic and emotional charge of the motion. Although all the temporal alignment methods described above have given satisfying results on motion sequences that are either cyclic (locomotion) or relatively short (martial arts moves), they are challenged when applied to long realizations of a CGS. The issue is actually well described by Keogh [27], who proposes an interesting discussion focused on the temporal aspects of human motion. He highlights the fact that temporal variability observed as both a local and a global influence.

His work introduces the need to deal with many different temporal variations levels. This leads us to design a new temporal alignment scheme that is multi-level.

Style robust representation space

It is difficult to formulate a precise definition of style as style is tightly mixed with captured motion data. An actor may perform a predefined motion sequence according to different moods, speeds, or expressivity clues, but, even when asked to be as neutral as possible, the actor will still convey his own *kinematic signature*. Still, each realization of a CGS will at least contain a common subpart that conveys the semantic of the CGS. Identification of this subpart motivates our investigations towards a low dimensional representation subspace for CGS. The construction of a style robust distance function is motivated by the assumption that the meaningful part of the gesture is embedded in the subset which presents the greatest variance.

Introducing Weighted PCA

The motion representation that we have designed to characterize the motion data in a reduced but still accurate orthogonal subspace is directly inspired from the work by Forbes & al. [26]. The weighted PCA-based representation has the advantage of providing a coarse to fine representation that is driven by the amount of variance observed by each principal component in the original space. Furthermore, the weighted scheme fits well with our re-

quirement: the meaningful content of an CGS is mainly driven by the actor's upper limbs. It thus makes sense to introduce a weighting scheme that highlights arm and hand motion. We thus consider both hands, upper body an lower body as four substructures of respective global weights $\{1.0, 1.0, 1.0, 0.5\}$. Inside a substructure, weights are derived from the relative amound of body mass influenced by each joint.

Motion data description

Motion data is composed by series of quite large vectors. As we deal with full body posture descriptions (hands + body), a posture vector is described by 63 unit quaternions. The quaternion representation is then centered and linearized thanks to the method presented in Johnson's PH.D thesis [28]. This prepossessing step leads to a linear real valued representation of our posture sequences described by a matrix M of 189 rows and n columns where n is the number of frames in the CGS realization. Among our database of motions, the mean frame number of a CGS realization is 7000.

Eigenposture base extraction

PCA is a linear basis transformation that basically decomposes the original data so that any number of components accounts for as much as possible of the data variance. Mathematically, the principal components are the eigenvectors of the covariance matrix of the original data set. To perform PCA decomposition, we rely on the singular value decomposition

(SVD) which, when applied to a reference CGS realization matrix M_{ref} leads to:

$$M_{ref} = U_{ref}\Sigma_{ref}V_{ref}^{T}.$$

Where V_{ref} and U_{ref} are orthogonal unit matrices, U_{ref} is an orthonormal eigenposture base of $\mathbb{R}^{189}$ whose r first columns give the basis $u1, u2, ...$ of the optimal hyperplane of dimension r. Σ_{ref} is a $189 \times n$ matrix with non-negative decreasing singular values on its diagonal. Then, the M_i subsequent realization of a CGS are projected onto the optimal basis extracted from M_{ref}.

$$V_i^T = \Sigma_{ref}^{+}U_{ref}^{T}M_i$$

Where Σ_{ref}^{+} is the transpose of Σ_{ref} with every nonzero entry replaced by its reciprocal. This projection leads to a common representation space between every realization of a CGS. In this representation space, a realization M_i of a CGS is described by the r first rows of matrix V_i. The distance between two poses is obtained by calculating the euclidian distance between their first r scaled coordinates of V_i.

Projecting the motion onto an optimal subspace

Projecting an arbitrary motion on an eigenposture space is lossy if the projected motion is not included in the construction of the eigenposture basis. To minimize the error induced by the projection, Forbes & al [26]. constructed the motion search space on an eigenposture

basis obtained from a motion sequence that conveyed the most variability: the range of motion (ROM) that had been used to calibrate the motion capture hardware. Such a choice is legitimate when no a-priori knowledge is available on the motion data to be projected.

Or we have a strong a-priori on the content of the motion data we deal with: each motion clip is a single realization of a CGS. As a consequence, it makes sense to build the projecting space upon a reference CGS realization that closely matches the motions we wish to compare rather than an exhaustive ROM. This decomposition leads to a more accurate representation space. In other words, fewer eigenpostures will be necessary to provide an accurate reconstruction of the projected motion. Therefore, fewer coordinates are required to provide a faithful estimation of the *fundamental motion*. In our experiments, we found that taking $r = 4$, was sufficient to convince the overall meaning of any realization of a CGS. We thus define our distance function δ between two postures q_i and c_j belonging to two coordinates matrices Q and C expressed in the reference base $(U_{ref}\Sigma_{ref})$ as:

$$\delta\{q_i, c_j\} = \sqrt{(q_i(1) - c_j(1))^2 + ... + (q_i(4) - c_j(4))^2}$$

This distance function serves as a base to the construction of a distance matrix that will be handled by the adaptive DTW algorithm we introduce in the next section.

Time alignment

Accurate matching of human motion data requires taking multiple levels of temporal misalignment into account, from uniform scaling along large sequences to small local misalignments. This fact has also been identified as well in biomechanical and computer animation [27]. To handle this issue, we rely on a multilevel strategy that iteratively adjusts both the search space and the slope constraint of the well known DTW algorithm [21]. By doing so, it becomes possible to avoid the trade-off between global and local adjustments. We show that this methods prevents discontinuous jumps from occurring while preserving sufficient accuracy in time correspondence.

Fast DTW

FastDTW algorithm has been introduced by Salvador & al. [29] and was initially designed to cut-off the computational cost of the well known DTW [21], which is of n^2 in its standard implementation. FastDTW basically consists in splitting the complexity of standard DTW by recursively down-sampling the time series. The warp path found at each iteration of the algorithm is then projected onto the higher resolution layer and serves as a guide that reduces computational complexity by spatially reducing the area handled by dynamic programming, as illustrated in figure 2b, FastDTW complexity is $O(n)$, and is known to find an accurate minimum-distance warp path between two time series that is nearly optimal. Unfortunately, the warp path obtained via FastDTW contains many consecutive horizontal or vertical steps.

This leads to jerkiness, high discontinuities or long steady postures over warped gesture sequences.

Constrained DTW

The constrained DTW has been introduced to limit the number of consecutive horizontal or vertical steps and provides a smoother match. Let Q and C be two time series of respective length m and n. Let δ be a distance function between any element q_i of Q and any element c_j of C. Let $k > 1$ be a real coefficient. Constrained DTW can then be recursively defined as:

$$\gamma\{q_i, c_j\} = \delta\{q_i, c_j\} + min(\gamma\{q_{i-1}, c_{j-1}\}, k\gamma\{q_{i-1}, c_j\}, k\gamma\{q_i, c_{j-1})\})$$

$$DTW(Q, C) = \gamma\{q_m, c_n\}$$

k prevents the warp path from leaving the main diagonal of the distance matrix between Q and C. Constrained DTW limits the number of consecutive horizontal or vertical steps and provides a smoother match. The drawback of constrained DTW is the slope limitation which is introduced. Slope limitation may prevent constrained DTW from finding a warp path that is faithful to the optimal warp path if the temporal variation has a too wide influence.

Adaptive DTW

Adaptive time warping was motivated by the wish to provide a constrained version of the DTW algorithm which could be adapted to the many temporal misalignment levels observed in the realizations of CGS. Existing time warping methods do not handle the trade trade-off between warp path smoothness and minimization of the warp path distance. To answer this limitation, a coarse to fine approach is adopted. This approach relies on a multilevel strategy that iteratively adjusts both the search space and the slope constraints of DTW. We thus extend FastDTW algoritm in order to take an adaptive slope support at each iteration into account. The pseudo code in figure 2a describes the algorithm.

The implementation is recursive and the base case occurs when one of the input motions becomes shorter than the window length we set in the parameters of Adaptive DTW. First, two new lower-resolution motion series are created that have half as many poses as the input motion series (Fig. 1, l. 5–6) Slope constraint is then updated (l. 7–9). Next, a low resolution path $[p, q]$ is found for the coarsened motion series (l. 10) and projected to a higher resolution (l. 11). This projected path is then expanded by radius cells to create a search window that will be passed to the constrained DTW algorithm (Fig. 1, l. 12 and Fig. 2). The constrained DTW algorithm refines the warp path that was projected form the lower resolution. The result of this refinement is then returned. The execution of the AdaptiveDTW algorithm repeatedly runs lines 5–9 in recursive calls to lower resolutions. When the base case is reached, it is executed only once, afterwards lines 11–12 are executed

for each recursive call (or resolution) on the stack.

Application

Communicative Gesture Sequence description

The earliest part of our work was dedicated to collecting sign language motion data performed according to various styles.

We then asked a professional signer to perform several realizations of the same weather forecast presentation in French sign language (FSL) by adopting different styles. The mean duration of a sequence was 60 seconds and the CGS realizations were done according to the following style: neutral (twice), emphasis, angry and tired. We took as reference realization the first sequence performed according to neutral style.

Results

We preprocessed and merged the data to obtain one unit quaternion posture matrix per CGS realization. We then projected the motions on the eigenposture basis extracted from the reference CGS realization. Finally, we applied standard DTW and Adaptive DTW to the obtained matrices. Figures 3 and 4 depict the most evocative results.

Figure 3 compares a temporal alignment obtained thanks to classical DTW, constrained DTW and the Adaptive DTW. The plots show the evolution of the influence of the first eigen-

posture vector for each motion in the eigenposture representation space. The two motions are respectively the first neutral sequence and the second neutral sequence.

Those sequences actually provide an interesting benchmark. On one hand, the sequences are very close, since they were performed according to the same style. On the other hand, they suffer from heavy artifacts since the signer messed up a couple of signs between frame 2200 and frame 3800.

The second plot from the top of figure 3 highlights the temporal alignment found by classical DTW. It appears that classical DTW actually finds an optimal time warp path which minimizes DTW cumulative distance. Unfortunately, the obtained warp path introduces many discontinuities, especially in the disordered part of the sequences. As a consequence, the warped motion obtained thanks to DTW contains many discontinuities.

The third plot from the top depicts the temporal alignment obtained thanks to constrained DTW. For the experiment, the maximum number of consecutive horizontal or vertical steps has been set to 2, as described in [24]. Although the warped motion presents satisfying continuity, It appears that constrained DTW is challenged when the temporal variations have a too wide influence over the motion to be warped. Adaptive DTW shows its ability to find a smooth but still accurate warp path between the two motions.

Figure 4 illustrates the time alignment obtained by aligning the angry styled CGS performance onto the reference CGS performance. On one hand, the spacial variations introduced by angry style are not negligible (notice the differences between the overall postures and the amplitude of the movements). On the other hand, angry style introduces rhytmic repetitions

of some specific movements which ADTW is not designed to handle. The postures depicted in figure 4 are equally sampled and the curves represent the evolution of the influence of the first the eigenposture vectors along frames. This figure highlights the capability of Adaptive DTW to provide a smooth and accurate match despite the spacial variations and repetitions introduced by the angry style.

Conclusion

Using communication gestures obtained through motion capture raises a number of difficulties that are inherent to the nature of these motions: difficult automatic or manual segmentation and strong variability between the execution of two realizations of a communicative gesture sequence (CGS). In this paper we have proposed a motion alignment method that has proved to be robust to the spacial variabilities that are induced by differently styled realizations of a relatively long CGS. This information can be exploited in a variety of ways: motion editing, blending, segmentation or style translation. Our hypothesis lies on a possible decomposition of communicative gesture between a fundamental motion (common to all realizations) and a style content. This decomposition is obtained through weighted PCA decomposition methods. This decomposition is then used as input of an adaptive dynamic time warping algorithm which provides smooth alignment between sequences. Style may introduce rhytmic repetitions of some specific movements. ADTW is not designed to handle such repetitions, but answering this limitation contitutes a challenging extention.

References

[1] S. L. Prillwitz, R. Leven, H. Zienert, R. Zienert, T.Hanke, and J. Henning. *HamNoSys. Version 2.0.* International Studies on Sign Language and Communication of the Deaf, 1989.

[2] T. Lebourque, S. Gibet, and P.F. Marteau. High level specification and animation of communicative gestures. *Journal of Visual Languages and Computing*, 12(12):657–687, 2001.

[3] I. Zwiterslood, M. Verlinden, J. Ros, and S. van der Schoot. Synthetic signing for the deaf: Esign. In *Proc. of the Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment*, Granada, Spain, July 2004.

[4] R. Laban. *The Mastery of Movement*. Northcote House, 1988.

[5] Diane M. Chi, Monica Costa, Liwei Zhao, and Norman I. Badler. The EMOTE model for effort and shape. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proc.*, pages 173–182. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[6] H.G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28:879–896, 1998.

[7] C. Pelachaud B. Hartmann, M. Mancini. Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture in Human-Computer Interaction and Simulation, LNAI*, 2006.

[8] M. Unuma, K. Anjyo, and R. Takeuchi. Fourier principles for emotion-based human figure animation. In *SIGGRAPH 95: Proc. of the 22nd annual conference on Computer graphics and interactive techniques*, pages 91–96, New York, NY, USA, 1995. ACM Press.

[9] A. Bruderlin and L. Williams. Motion signal processing. In *SIGGRAPH 95: Proc. of the 22nd annual conference on Computer graphics and interactive techniques*, pages 97–104, New York, NY, USA, 1995. ACM Press.

[10] K. Amaya, A. Bruderlin, and T. Calvert. Emotion from motion. In Wayne A. Davis and Richard Bartels, editors, *Graphics Interface 96*, pages 222–229. Canadian Human-Computer Communications Society, 1996.

[11] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. In *Proc. of Int. Conf. on Computer Graphics and Interactive Techniques*, pages 473–482, San Antonio, USA, 2002.

[12] J. Lee, J. Chai, P. Reitsma, J. K. Hodgins, and N. Pollard. Interactive control of avatars animated with human motion data. In *Proceedings of SIGGRAPH 2002*, 2002.

[13] O. Arikan, D. A. Forsyth, and J. F. O'Brien. Motion synthesis from annotations. *ACM Trans. Graph.*, 22(3):402–408, 2003.

[14] Y. Cao, P. Faloutsos, and F. Pighin. Unsupervised learning for speech motion editing. In *Proc. of ACM SIGGRAPH/Eurographics, SCA*, pages 225–231, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[15] A. Shapiro and P. Faloutsos. Style components. In *Proc. of Graphics interface*, 2006.

[16] R. Boulic P. Glardon and D. Thalmann. Pca-based walking engine using motion capture data. In *In Proceeding of Computer Graphics International (CGI)*, pages 292–298, 2004.

[17] C. Bregler E. Chuang, H. Deshpande. Facial expression space learning. In *Proc. Pacific Graphics*, 2002.

[18] D. Vlasic, M. Brand, H. Pfister, and J. Popović. Face transfer with multilinear models. *ACM Trans. Graph.*, 24(3):426–433, 2005.

[19] M. Brand and A. Hertzmann. Style machines. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 183–192. ACM Press / Addison Wesley Longman, 2000.

[20] H.-Y. Shum Y. Li, T. Wang. Motion texture: a two-level statistical model for character motion synthesis. In *proc. of Siggraph 2002*, pages 465–472, 2002.

[21] C. S. Myers and L. R. Rabiner. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal*, 1981.

[22] A. Witkin and Z. Popović. Motion warping. *Computer Graphics*, 29:105–108, 1995.

[23] C. Rose, M. Cohen, and B. Bodenheimer. Verbs and adverbs: Multidimensional motion interpolation. *IEEE Computer Graphics and Applications*, 18:pp. 32–40, 1998.

[24] L. KOVAR and M. GLEICHER. Flexible automatic motion blending with registration curves, 2003.

[25] E. Hsu, K. Pulli, and F. Popović. Style translation for human motion. *ACM Trans. Graph.*, 24(3):1082–1089, 2005.

[26] K. Forbes and E. Fiume. An efficient search algorithm for motion data using weighted pca. In *SCA '05: Proc. of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 67–76, 2005.

[27] A. Wai-Chee Fu, E. keogh, L. Yung Hang Lau, and C. A. Ratanamahatana. scaling and time warping in time series querying. In *vldb '05: Proc. of the 31st international conference on very large data bases*, pages 649–660. vldb endowment, 2005.

[28] M. P. Johnson. *Exploiting Quaternions to Support Expressive Interactive Character Motion*. PhD thesis, Massachusettes Institute of Technology, 2003.

[29] S. Salvador and P. Chan. Fastdtw: Toward accurate dynamic time warping in linear time and space. In *KDD Workshop on Mining Temporal and Sequential Data*, pages 70–80, 2004.

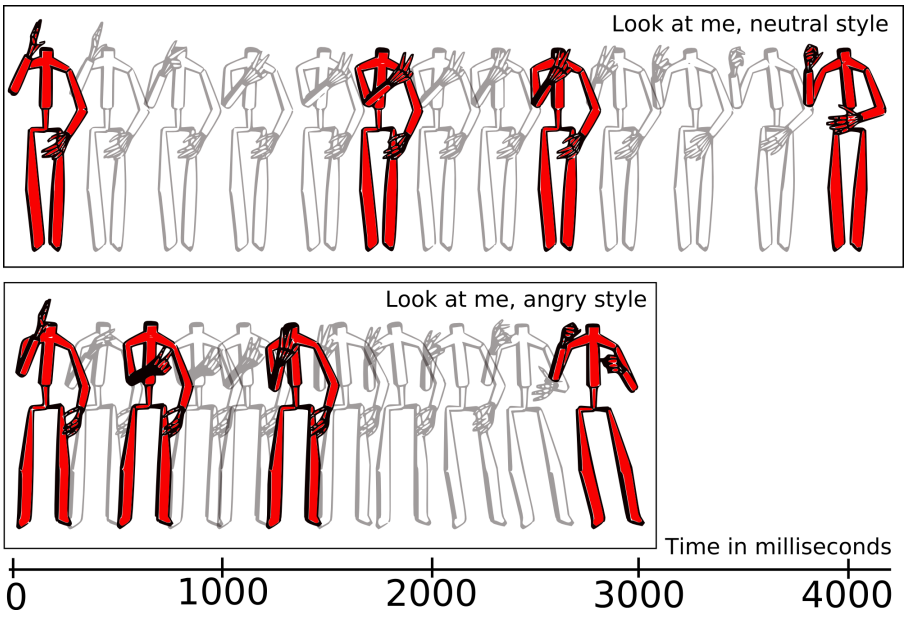

Figure 1: FSL signer performing the sign ”look at me” on the top sequence, he was asked to be as neutral as possible, on the bottom one, he was asked to simulate angryness. The two sequences are displayed along the same time basis

(a)

AdaptiveDTW(X,Y,radius,K)

Require:

X - a projected motion

Y - a projected motion

radius - distance to search out of the projected warp path from the previous resolution

K - slope constraint at the finest resolution

Ensure:

[p,q] a warp path along the motions X and Y

```
1: minTSsize = radius + 2 \\ The min size of the coarsest resolution
2: if length(X) ≤ minTSsize OR length(Y) ≤ minTSsize then
3:    RETURN DTW(X,Y,K)
4: else
5:    shrunkX = downsample(X) \\ length(shrunkX) = 1/2 × length(X)
6:    shrunkY = downsample(Y)
7:    if K > 1 then
8:       K = K − 1 \\ Adjust slope constraint for coarser resolution
9:    end if
10:   (p,q) = AdaptiveDTW(shrunkX,shrunkY,radius,K)
11:   window = ExpandResWindow(p,q,X,Y,radius)
12:   RETURN DTW(X,Y,window,K)
13: end if
```

(b)

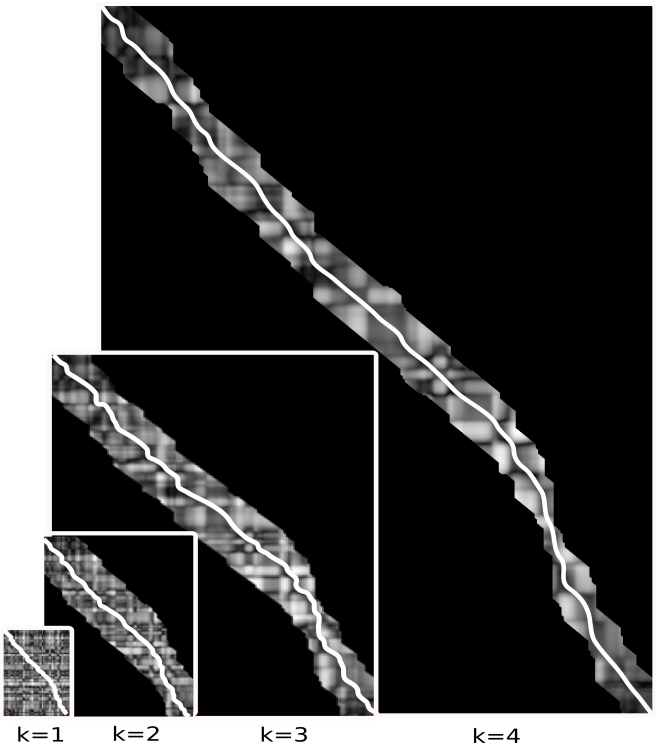

Figure 2: (a) The Adaptative DTW algorithm. (b) At each iteration, as the Warp path is projected onto the higher resolution, the slope constraint of the DTW algorithm is incremented

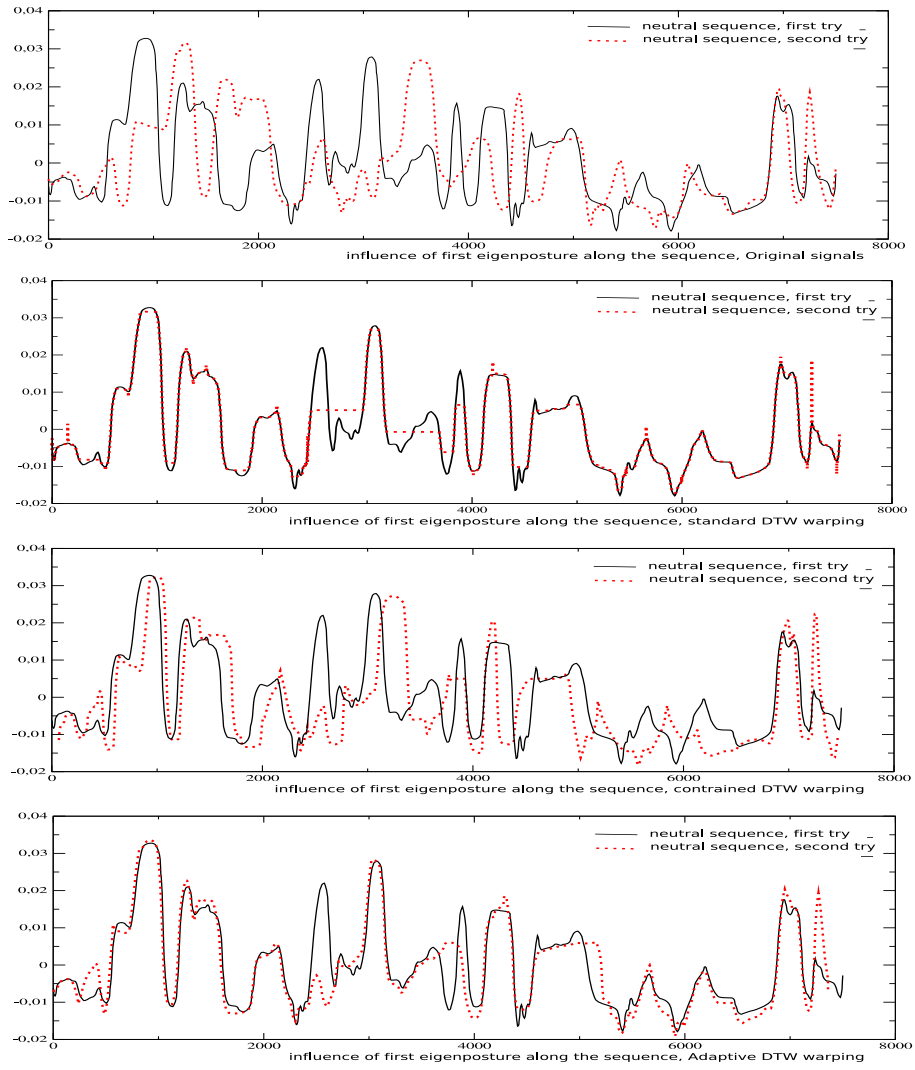

Figure 3: Comparison between standard DTW algorithm, constrained DTW and Adaptive DTW (bottom): ADTW does not find an temporal alignment that minimises DTW distance, but, rather a smooth path that prevents jerkiness along the warped motion. moreover, ADTW recovers both global and local time variations.

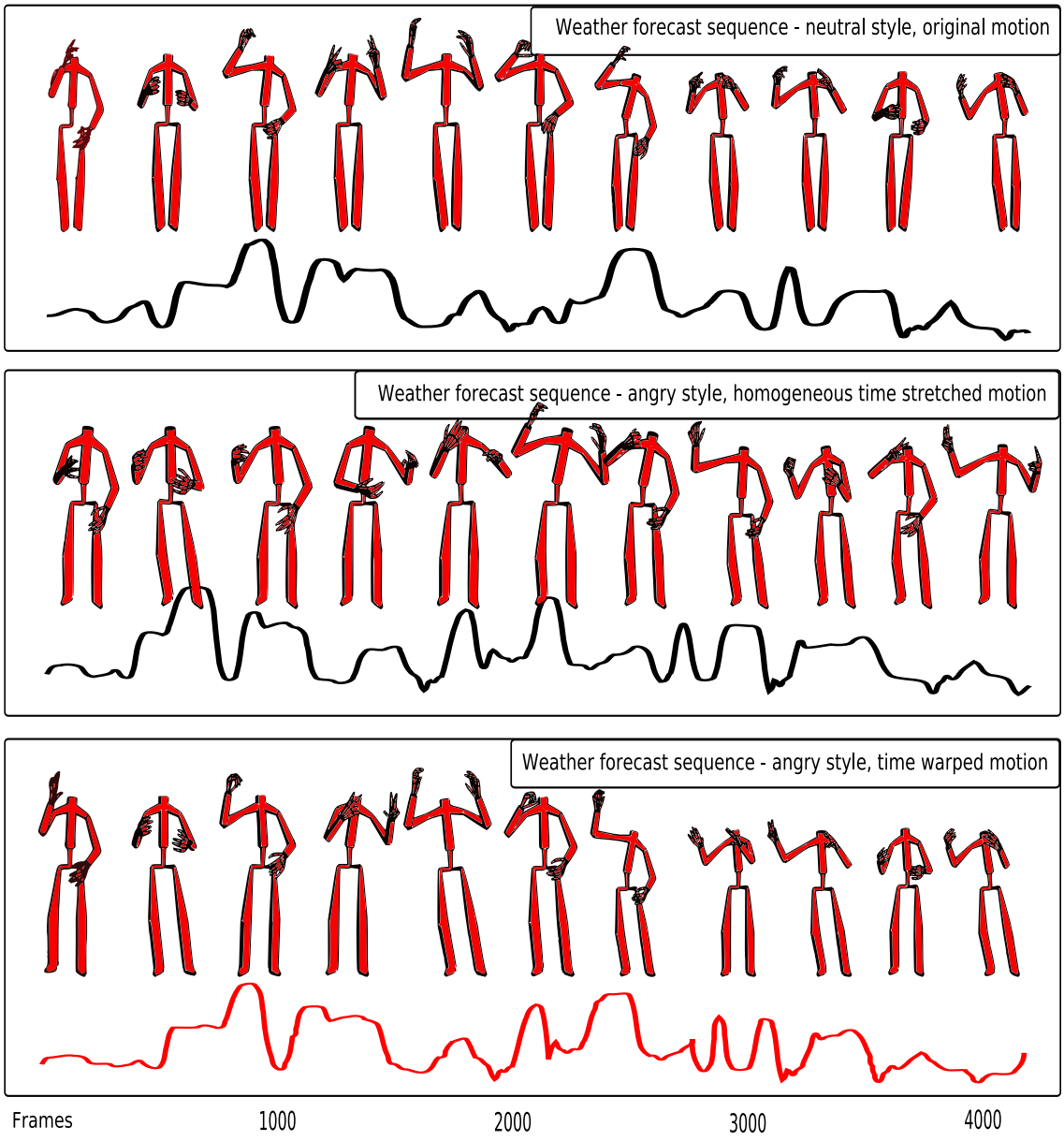

Figure 4: Pose representation along three sign sequences. From top to bottom. The two upper sequences are captured from original performances according to neutral and angry style, respectively. The third sequence represent the original angry sequence warped along the original neutral sequence.